DOCUMENT RESUME

ED 155 218 TM 007 347

AUTHOR Wood, Peter H.
TITLE Student and Peer Ratings of College Teaching and Peer Ratings of Research and Service: Four Years of Departmental Evaluation.
PUB DATE [Mar 78]
NOTE 59p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS Bias; College Faculty; *College Teachers; *Comparative Analysis; *Evaluation Criteria; Evaluation Methods; Evaluators; Higher Education; *Peer Evaluation; Peer Relationship; Reliability; *Student Evaluation of Teacher Performance; *Teacher Evaluation

ABSTRACT

Faculty of the education department of a midwestern university were rated as teachers by both students and peers, and ranked by peers as to their effectiveness in service and research. Student ratings were gathered from three questions on The Student Description of Teaching form, which concerned assessment of the teacher, the is appended. Access to personnel files was provided when colleagues ranked total course experience, and the students' accomplishment in the course. A sample form is appended. Access to personnel files was provided when colleagues ranked departmental faculty members from first to seventh as teachers, researchers, and providers of service. Peer evaluations were also obtained from a Faculty Evaluation Committee which rated departmental faculty members on the basis of their vitae. Four years of rating/ranking indicated that: student ratings were quite reliable and quite stable from year to year, but usually unrelated to peer rankings of teaching. Peer evaluations of teaching, research, and service were similarly stable from year to year; but peer evaluations of research were significantly related to, and possibly biased by, inter-peer association. Peer evaluations of teaching were related both to peer-association and to office location or faculty visibility. Peer evaluation of service was strongly related to committee memberships but relatively unrelated to a faculty association measure.
(Author/JAG)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED155218

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Peter H. Wood

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Student and Peer Ratings of College Teaching and Peer Ratings of Research and Service: Four Years of Departmental Evaluation

Peter H. Wood
Bowling Green State University*

Running Head: A Four Year Evaluation of a College Department

*Peter H. Wood
#510, 414 West 120th Street
New York City, New York 10027
Printed in the U.S.A.

IM007 347

Student and Peer Ratings of College Teaching and Peer Ratings of Research and Service: Four Years of Departmental Evaluation

PETER H. WOOD, Bowling Green State University

Abstract

Faculty of a 26-person college department of a midwestern university were rated and/or ranked as teachers by both students and peers, and also ranked by peers as to their effectiveness in service and research. Four years of rating/ranking indicated that: student ratings were quite reliable (~.92) and quite stable from year to year (~.68) but usually unrelated to peer rankings of teaching. Peer evaluations of teaching, research and service were similarly stable from year to year, but peer evaluations of research were significantly related to (biased by?) inter-peer association while peer evaluations of teaching were related both to peer-association and to office location or faculty visibility. Peer evaluation of service was strongly related to committee membership but relatively unrelated to a faculty association measure.

Student and Peer Ratings of College Teaching and Peer Ratings of Research and Service: Four Years of Departmental Evaluation

Background - Perspectives

General. Sharply reduced faculty mobility (Cartter, 1976, p. 161), increased demands for educational accountability, active enforcement of antibias regulations (Lester, 1974), and increasingly rigorous promotion and tenure procedures (Shull, 1976) have created an increased need to clarify procedures for faculty evaluation. The majority of college deans surveyed by Sedin and Wakin (1974, pp. 48-49) were critical of their own evaluation procedures and admitted that their personnel decisions were not based upon "rational, impersonal, and unprejudiced information" (p. 49). Comparisons between the results of the Sedin and Wakin survey of 1974 and the Astin and Lee survey of 1966 do, however, indicate a trend toward the increased employment of more structured and systematic faculty evaluation procedures which are more likely to produce outcomes that are open to rational and empirical examination. The use of formal and relatively systematic student ratings of teaching had surpassed the gathering of more informal student opinions; committee evaluation of faculty performance had increased while the dependence upon informally obtained collegial opinion had declined; and research indicators had become less frequently used as measures of teaching effectiveness.

Faculty evaluation procedures serve functions other than those related to the making of re-appointment, promotion, tenure, and salary decisions. The recent and widespread creation of college and university "faculty development centers" has been documented by John Centra (1976). Responses to his survey demonstrated that relatively systematic, faculty

assessment by students and colleagues is becoming an increasingly common component of faculty evaluation for purposes of faculty improvement.

Local background. In the past several years, the practical implications of national trends have been brought to the attention of administrators at Bowling Green State University (BGSU). BGSU administrators: (a) have been sued by faculty or prospective faculty who claimed unfair personnel procedures; (b) have been pressured to re-allocate faculty positions due to changes in patterns of student enrollment; (c) have had to construct, but not yet implement, contingency plans to meet potential reductions in student enrollment; and (d) have been required by their Board of Trustees to institute merit systems which allocate a significant segment of any annual salary increases as a reward to especially "meritorious" faculty --- and to other meritorious personnel.

The EDFI Department. Approximately six years ago, the BGSU Department of Education was subdivided into four smaller departments of from twenty to thirty faculty each. One of these four departments, Educational Foundations and Inquiry (EDFI) created a Faculty Evaluation Committee (FEC) to advise the chairperson and the department concerning the personnel decisions of hiring, re-appointment, promotion, tenure, differential merit pay, and resolution of salary inequities. This committee was composed of five voting members: one elected biannually from each of the four faculty ranks (instructor, assistant professor, associate professor, professor); with the fifth member appointed annually by the department chairperson. The department chairperson and the departmental representative to the college personnel committee hold ex-officio, but non-voting, committee membership. Discomfort with purely

subjective evaluations of teaching, research, and service (the three faculty functions specified but not defined in the BGSU Faculty Charter, 1969, p. 27), and pressure to implement an objective and fair merit system led to the gathering of judgments concerning faculty effectiveness (merit) from those sources most readily available -- the faculty vita, and student and colleague judgments. The data presented in this paper resulted from the analysis of some of the outcomes of this evaluation system.

Faculty Evaluation: Previous Research

E.R. Guthrie (1949, 1954) and Swanson and Sisson (1971) have reported the results of two faculty evaluation systems. Both systems drew upon the same three sources of information: (a) faculty data sheets which were circulated to faculty peers and administrators; (b) faculty and administrative judgments as to faculty effectiveness; and (c) student ratings of classroom teaching. Swanson and Sisson (1971) asked each of the members of a department (of 12) to nominate the four faculty most meeting the ideals of teaching, research and service. The department chairperson was asked to rank all departmental members on this same single factor. Guthrie (1949) had ad hoc groups of seven colleagues (five fellow faculty, the department chairperson and, often, the college dean) rate candidates for personnel decisions on the nine factors most frequently mentioned as items that faculty believed should be considered for promotion. Teaching, research, departmental and campus activity, and six less frequently mentioned factors had been elicited by a faculty questionnaire before Guthrie (1949) constructed his evaluation system. The University of Washington student rating form (French, 1957) was employed by Guthrie as a measure of student rating of

teaching effectiveness while Swanson and Sisson (1971) employed the CEQ form developed at the University of Illinois (Spencer, 1969a, 1969b).

Three outcomes were associated with both studies, and have been also supported by the majority of other research in faculty evaluation. Firstly, student ratings of teaching are quite reliable and reasonably stable. Guthrie (1949) reported that ratings of groups of twenty students correlated well (r = .87 and .89) with ratings by similar groups of students. The CEQ form used by Swanson and Sisson has produced internal-consistency reliabilities exceeding .90 (Spencer, 1969a, 1969b; Swanson and Sisson, 1971). In spite of extensive and common theoretical and methodological defects of design (Sockloff, 1973), the reliabilities of student ratings (by classes of twenty or more students) is comparable to that of professionally constructed achievement measures (Centra, 1972; Costin, Greenough and Menges, 1971; Doyle, 1975). More recent, empirical research (Bausell, Schwartz and Purohit, 1975; Hogan, 1973) and generalizability investigations (Gillmore, Kane and Naccarato, 1977) have provided evidence that sets of teachers, teaching the same course in subsequent terms or years, produce correlations in the .69 to .80 range, while producing consistency coefficients in the .33 to .48 range if teaching different courses. Although Swanson and Sisson (1971) did not provide similar consistency data, an analysis of their rank-order data (Tables 8 and 9) produces consistency coefficients (across adjacent quarters) of .44, .72 and .99 for student, peer and chairman rankings. However, Centra (1976) and Centra and Creech (1976) have provided data indicating that inexperienced teachers may elicit lower and more variable student ratings. Centra and Creech (1976) have also

identified some of the elements -- class size, level and type; expected grade; and teacher experience -- that often influence, and possibly, bias student ratings and that may reduce ratings stability.

Secondly, peer judgments also can be quite reliable. Guthrie (1949, 1954) reported that groups of seven colleagues produced judgments of nine faculty functions that correlated well (r = .64 to .76) with the judgments of different juries of seven. Similar stability coefficients calculated from the Sisson-Swanson data, but from nominations by the same colleagues only three months later, were .72 for peers and .99 for the chairperson. Peer nomination systems in other settings have produced internal-consistency coefficients ranging up to .90 (Hollander, 1956; Wolin, 1956) and have demonstrated considerable predictive validity for a range of academic and career outcomes in spite of their possible contamination by "friendship" factors (Hollander, 1956; Wherry and Fryer, 1945).

Thirdly, student and peer judgments of teaching have been positively correlated, although the typically moderate degree of correlation indicates that there were decided differences between the two types of judgments. Reported correlations are: .48 (Guthrie, 1949); .69 (Maslow and Zimmerman, 1956); .49 (Costin, 1966); .62 (Blackburn and Clark, 1971); .87 (Murray, 1972); .27 and .28 (Aleamoni and Yimer, 1973); .11 to .49 (Braunstein and Benston, 1973); and .84 (Ballard, Reardon and Nelson, 1976). The Swanson-Sisson correlations were between faculty nominations of total (teaching/research/service) effectiveness vs. student ratings of teaching, and the two rank-order correlations were .70 for the winter quarter and .28 for the following quarter (spring).

Two recent studies involving classroom observers (Centra, 1975; Love, Minnis, Sandoval and Yonge, 1977) also support the position that student members of a class and "outsiders" perceive teachers and teacher functions differently. The results of a study in a military setting (Morsh, Burgess and Smith, 1956) included a high correlation between supervisor and peer-instructor rankings (.77), moderate correlations between student ratings and learning gains (.41 to .46), but negligible correlations between supervisor or peer rankings and student gains attributable to instructors. Another potential bias in peer judgments emerged in the Hildebrand, Wilson and Dienst (1971) discovery that faculty included research and service descriptors in their definitions of the "best" vs. the "worst" teachers.

With a few exceptions, most researchers from Guthrie on have produced data to support the position that research and teaching effectiveness tend to be relatively independent faculty functions (Faia, 1976; Hoyt and Spangler, 1976), although a research reputation may moderate peer or administrator judgments concerning teaching effectiveness. The wide range of valued faculty functions elicited by Guthrie's (1949) questionnaire, the differently described functions of education faculty identified by Clark and Guba (1976), and the multiple definitions of the faculty service function presented in Martin's Redefining Service, Research and Teaching (1977) all suggest that the boundaries separating research and service are not defined precisely. Some of the Clark and Guba (1976) data could be interpreted as indicating that research and service functions may become especially blurred in departments or colleges characterized by low levels of knowledge production. Research and discussion presented by Smart and Montgomery (1976) add an

additional complication by calling attention to the diversity of processes, values and products that are associated with affiliation in different types of university departments. Consequently, results derived from one type of department may not generalize to a different type of department, and different departments may define teaching, research and service in different ways.

Objectives of this Study

The principle objectives of this study were: (a) to determine the cross-year consistencies of student ratings of the teachers of a moderately sized college department; (b) to determine the similar consistencies of peer rankings of the same teachers; (c) to investigate the relationships existing between student and peer judgments of teaching, and among peer judgments of teaching, research and service; and (d) to examine the relationships existing among peer judgments of teaching, research, service, social/academic association, and other characteristics which might influence peer judgments -- sex, rank, years of service, subdepartmental area affiliation, office location, and counts of typical research and service activities.

Although this study was intended to be a formative analysis of an ongoing evaluation system, some directional hypotheses seem justified because of similarities between this study and previous research.

1. Student ratings, for experienced faculty, will be correlated positively with the student ratings attained by the same faculty in adjacent years.

2. Student and peer evaluations concerning the teaching effectiveness of experienced teachers will be positively correlated.
3. Student ratings of the teaching effectiveness of experienced teachers will not be correlated with peer evaluations of research or service.
4. Student ratings of the teaching effectiveness of experienced teachers will not be correlated with counts of research or service activities.
5. Peer evaluations of research and service will be positively correlated.
6. to 8. A peer indicator of social/academic association will be positively correlated with peer evaluations of: (6) teaching; (7) research; and (8) service.

Method

Setting

This study focused upon the faculty of the EDFI Department, one of ten departments grouped within the College of Education, which is one of the six colleges or schools of a midwestern university. The university has been limited to approximately 15,000 students by the state legislature, and approximately 4,000 of these students are enrolled within the College of Education. No specific faculty evaluation system is mandated by either the university or the College of Education, although a series of general evaluation procedures have been in the process of development for the past five years, and some type of procedure for the

identification of meritorious faculty has been required of each department, every spring quarter, since 1974. Although candidates for promotion or tenure must provide some evidence of student evaluations for three academic quarters of teaching, no particular procedure is mandated by the college or the university. The EDFI department has required the use of the same student rating form, for every on-campus class, for the past four years.

The EDFI Department. During the period of this study, the EDFI Department was composed of from 28 to 29 faculty, 20 of whom taught classes during the four years of this survey and also had been teaching for at least two previous years, and 26 who are represented in some of the data.

An unpublished survey by the author of this study indicated that the mean percentage of faculty time devoted to teaching, research and service was: 62%, 17% and 19.5%, respectively. The preferred time allocation was 54% teaching, 30% research and 17.5% service. Approximately 82% of the faculty considered teaching to be "centrally important," 17% considered research to be "centrally important," and none so rated service. The majority (68%) believed that both teaching and research would increase in importance over the next five years, and 75% judged that both teaching and research were essential for tenure and promotion -- but that only teaching was supported actively by the department, college, or university.

Previous experience as a pre-college teacher was a departmental hiring pre-requisite, and an earned doctorate was a pre-requisite for tenure or rank beyond the instructor level. Of the 26 faculty repre-

sented in this study: 9 were associated with the educational/developmental psychology area; 4 were associated with the history/philosophy of education; 4 with guidance and counseling; 8 with tests and measurement; and 1 had dual association. Year of entry to the university ranged from 1955 to 1974 with a median year of 1968. Although faculty rank changed over time, during 1976, there were 3 instructors, 5 assistant professors, 9 associate professors, 7 full professors, and 2 people in transition. Thirteen faculty had offices on one of the two major corridors (4 on a central corridor dominated by the department, 9 on a corridor shared with another department), 4 had offices in a corridor on the same floor but dominated by another department, 6 had offices on another floor, and two had offices outside of the education building. The chairperson had his office adjacent to the central, departmental office which housed the secretary, student assistants, mailboxes, supplies and duplicating equipment.

Frank R. Kemerer has described some college or departmental cultures as "egalitarian" and resistant to permitting tenured professors to hold disproportionate influence in departmental decisions, or as more traditional, more established and more willing to believe that rank, length of service, research activity, and professional status entitle some faculty and administrators to exert substantial control over academic affairs (1978, p. 237). The EDFI department would be more accurately characterized as egalitarian than as traditional, and research productivity is as characteristic of the newer faculty as it is of the more experienced faculty, although service activity -- especially membership on university-level committees -- is more characteristic of the

more experienced faculty. Because of differing amounts of pre-college, teaching experience, full professors are not always older instructors.

Instruments and Procedures

Student ratings. Student ratings were defined as the total mean score resulting from all student responses to the three most general questions of the Student Description of Teaching form. Student ratings (1974) included all on-campus classes during the winter quarter of 1974. Student ratings (1975) included all on-campus, departmentally-assigned classes for three quarters -- Spring 1974, Fall 1974, Winter 1975. Student ratings 1976 and 1977 were similarly derived from all classes for three terms. The Student Description of Teaching was developed over a two year period to provide the department with data for personnel decisions while providing more detailed, optional feedback to each teacher. The first five questions were modifications of the Hildebrand, Wilson and Dienst (1971) five general ratings questions and were itended to orient the rater to five valued teacher characteristics. The following three questions, #6, 7, and 8, were used to produce the student ratings score used in personnel decisions. These three questions asked the student to produce a general assessment of: (a) the teacher; (b) the total course experience; and (c) their own accomplishment in the course. The remaining questions were used for research purposes and to provide more specific data to the teacher.

Reliability. Scale reliability was estimated during the Fall Quarter of 1974 by drawing 1,609 students from EDFI classes taught by 28 teachers. Reliability was determined by analyzing the variance associated with all students and from randomly drawn sets of 5, 5, 10, 15,

and 20 students taken from the largest class of each teacher (see Winer, 1962, pp. 124-132). Resultant reliability estimates for the mean of the three general questions for different numbers of students were: 5 (.72, .76); 10 (.88); 15 (.90); 20 (.92); all students (.92). Obviously, this indicator of internal consistency confounds student, course, term, and instructor effects. The mean score for 1,609 students was 3.9; range = 1 to 7; SD = 1.8; and skew = -.33.

Peer Evaluations. In the spring of 1974, all departmental faculty were asked to use three, five-point scales (1 = low, 5 = high) to rate all other faculty on the three dimensions of teaching, research and service. Ratings reliabilities were estimated via analysis of variance and were computed as .62, .63, and .63 for teaching, research and service. Only 20 of 26 potential raters completed the rating scale, and the members of the department complained that they could not honestly rate all members of the department since they were not knowledgeable concerning the activities of many colleagues.

During the Spring Quarters of 1975, 1976 and 1977, all faculty were provided access to all departmental personnel files and were asked to rank departmental faculty from first up to seventh according to the degree to which faculty members personified the active and effective teacher, researcher and provider of service. Several possible criteria for each function were presented at the top of each of the three ranking forms. Each of the three forms also listed the names of all departmental faculty. Eighteen of 26 faculty completed the ranking forms in 1975, and the reliability estimated for the teaching, research, and service dimensions were .78, .82 and .82 respectively. All but one or

two faculty completed the ranking forms for 1976 and 1977, but reliability estimates were not calculated because access to the raw data was restricted in order to preserve the complete anonymity of those providing the rankings. The two scores derived from the ranking procedures were: number of times ranked in the top seven; and the total summed raw score (failure to be ranked in the top seven earned an automatic value of 8). Correlations between these two possible scores ranged from -.96 to -.94. The raw score was chosen as the indicator of teacher rank.

Committee (FEC) Ratings, 1975. In 1975, the five members of the newly organized Faculty Evaluation Committee independently examined all faculty vitae and rated each departmental member on a five point scale on each of the three dimensions of teaching, research and service. The five ratings were summed to produce a score on each function for each faculty member. The estimated inter-rater reliabilities for these scores were .74, .80 and .84 for teaching, research and service. The department subsequently voted to eliminate this procedure. Consequently, there were no FEC ratings for 1976 and 1977.

Visibility. Each faculty member was categorized as to potential visibility to other faculty. Faculty were assigned a visibility label associated with their office location and its relationship to the departmental office and to the density of other departmental members.

Rank, Area and Enter. All faculty were categorized as to professorial rank, departmental area subdivision and according to the year they began teaching at BGSU.

Activity Counts. The vitae, as of the end of the 1976-1977 academic year, of 26 faculty were examined to produce the following activity counts. Any activities previous to 1974 were not counted. 1) Committee membership = total number of years of service on all listed departmental, college, university or ad hoc committees. Each year of service on any committee earned one point. Committee chairship, for any period of time, earned an additional two points. 2) Books, = cited authorship or co-authorship of any book or book-like collection. 3) Articles = the sum of all cited articles published in any sort of journal or journal-like outlet. 4) Papers = sum of papers read before professional associations, research reports presented to any authority and published book reviews. 4) Workshops/speeches = the sum of all workshops or speeches presented to non-professional or other, potentially non-critical, audiences. 6) Grants = all local or national grants received for any education related purposes.

Association. Twenty-six departmental faculty were presented with a list of the 26 faculty represented in the data and asked to rank (1 = most closely associated) the degree to which they associated with, socially and/or academically, the listed faculty during the 1976-1977 academic year ending in June 1977. The total number of rankings in the top seven ranks represents the association score of each faculty member.

Teaching Assignment Stability. The courses taught (as part of the regular, on-campus schedule) for each of the last three rating years (the 1975, 1976, and 1977 rating years each began in the Spring term of the preceding year and continued through the Fall and Winter Quarters) were tallied. Four stability scores were created: (a) Courses =

number of separate sections; (b) Graduate = number of graduate classes or sections; (c) Small = number of courses or classes or sections smaller than 16 students; and (d) New = number of previously untaught classes.

Advisement. Four advisement counts were derived from the vita: (a) average number of undergraduate advisees; (b) average number of graduate advisees; (c) total number of theses committees for the last five years; and (d) the sum of the previous three counts.

Course Load. Four course load indicators were created from four years of departmental records: (a) Classes = total number of different courses taught in a four year period; (b) SCH = total number of student credit hours assigned to an instructor over a four year period; (c) Extra = student credit hours gained by teaching non-assigned classes such as those courses offered by the experimental studies or university division programs, and independent study or thesis research or readings classes; and (d) Regular = total student credit hours associated with regularly assigned, on-campus classes.

Total Activity Count. Since faculty evaluation of other faculty may be influenced by an entire history of research or service rather than just current research or service activity, five activity counts were drawn from the faculty vita. Although the previously-mentioned activity counts only included activities performed during or after 1974, these counts included all activities mentioned in the vita and accomplished between arrival at BGSU and submission of the vita in the late Spring of 1977. These four total counts were: (a) Total articles = the sum of all articles reported as published in any type of journal

or journal-like publication; (b) Papers-Grants = the total of all published book reviews, research reports, grants received or papers presented to professional groups; (c) Speeches-Workshops = the sum total of all speeches or workshops reported in the vita; and (d) Consultantships = all consultantships reported in the vita. The number of new courses or programs created by a faculty member was reported in the vita for some faculty but not for others and was not counted, although it might possibly influence faculty evaluation of teaching or service. These total activity counts were added to the study after the first few data analyses indicated that the four-year activity counts were poor predictors of peer ranking of research.

Results

Cross-Year Consistency of Student Ratings

Student ratings were collected for the four years ending in the mid-spring of 1974, 1975, 1976 and 1977. Although the 1974 ratings only represent one academic quarter, the Winter Quarter of 1974, and may only include one or two classes for the faculty who had voluntarily used the same ratings form, the ratings for 1975, 1976 and 1977 represent all (or almost all) on-campus classes taught during three academic quarters (Spring, Fall and Winter). The resulting yearly mean ratings, standard deviations and between-year correlations are presented in Table 1. The correlation coefficients above the main diagonal represent only the more experienced faculty who were hired before 1972, while the coefficients below the main diagonal represent all faculty. The year-to-year sta-

bilities of the more experienced faculty are .65, .71 and .81 for the 1974 to 1975, 1975 to 1976 and 1976 to 1977 periods. When the beginning teachers are added to the pool, the corresponding stabilities for the entire department are .49, .75 and .72. The stability across a two-year span (1974 to 1976 and 1975 to 1977) is .64 and .79 for experienced faculty and .46 and .45 when all faculty are considered. The three-year stability for experienced faculty is .54 for experienced faculty and .38 for all faculty. With one exception, the smallest stability coefficients all involve the truncated and limited 1974 ratings which only represented one academic quarter. At least some of the lower stability for all faculty seems to be attributable to the initial improvement and subsequent stabilization of the ratings produced by some of the newer faculty. A slight drop in departmental ratings for all faculty occurs in 1975 when six newer faculty are added to the pool, and this drop, from 4.25 to 4.13, is recovered in subsequent years.

The cross-year stabilities for experienced faculty for the three complete years of ratings gathering all exceed .7 and are all significant at the .001 level. The corresponding, single-year stabilities for all faculty are also above .7, however the 1975-1977, two-year stability for all faculty is only .45 and only significant at the .05 level.

Place Table 1 about here

Similarity of Student and Peer Evaluation of Teaching

Table 2 presents the means and standard deviations derived from student and peer evaluations of teaching. The student evaluation

scores, as in the previous section, are based upon the mean student response to three very general questions related to: (a) instructor effectiveness; (b) total class experience; and (c) student accomplishment. The possible student ratings range is from a low of 1 to a high of 7; the teacher score range was from 2.58 to 5.60; and the departmental range presented in Table 2, is from 4.13 to 4.36. The mean departmental responses are somewhat closer to scale point #4 (Better than most, Best 30%) than to scale point #5 (Superior, Top 10-15%). The student ratings scale is presented in the appendix to this paper as are the descriptors for the peer nominations for teaching, service and research.

The peer ratings or rankings are somewhat more difficult to interpret since three different systems are represented. The 1974 peer ratings are the mean score produced when all faculty were asked to rate all other faculty, as "teachers," on a five point scale with 5 as high and 1 as low. The resulting score range was 2.1 to 4.0 with a median score of 2.88. The mean peer ratings for 1974 were 2.8 for all faculty and 2.9 for experienced faculty.

The peer nomination-rankings for the successive years of 1975, 1976 and 1977 were all produced by asking faculty to nominate and rank-order up to seven faculty as the <u>best</u> teachers (or researchers or providers of service). Faculty not ranked in the top seven by a peer were automatically assigned a number of eight. Consequently, the best possible peer nomination-ranking score would be 1 times the number of participating faculty, if all faculty chose that person as first. The worst possible score would be assigned to any teachers not ranked in the top seven by

any other peer--including themselves. Score sizes would be larger when more faculty participated and smaller when fewer faculty participated. In 1975, 26 peers participated to produce a potential score range of 26 to 208, an actual score range of 151 to 208, a median score of 196 and a mean score of 189 for all teachers and 189 for experienced teachers. Thirty peers in 1976 produced a possible score range of 30 to 240, an actual range of 156 to 240, a median of 222 and a mean of 218 for all teachers and 220 for the more experienced teachers. The 1977 scores were: 24 to 192 possible range, 89 to 174 actual range, 160 median, 155 mean for all faculty and 157 mean for experienced faculty. Although faculty were permitted to use their own ranking criteria, seven potential criteria were provided for their consideration. One of the seven suggested criteria was similar to the "accomplishment" question of the student ratings form, and another, "knowledge of content area" was also suggested for student consideration, but the others related to teaching functions not as available to classroom students: sharing with colleagues; efforts to improve; advisement; supervision of theses and independent study; and development of innovative courses or programs.

The FEC 75 scores were the sum of the five scores produced by the five members of the 1975 Faculty Evaluation Committee, each using a 1 to 5 scale to produce a possible score range of 5 to 25, an actual range of 10 to 22, a median of 15.5, and means of 16.3 for all faculty and 16.7 for experienced faculty. The 1974 all-rate-all procedure was eliminated by departmental vote as requiring people to rate people they didn't know well, and the FEC ratings were also voted down as redundant and as providing too much power to a standing committee. The Summary Student

Ratings and the Summary Peer Rankings were each produced by adding the yearly mean scores for the last three years, 1975 to 1977, and dividing by three.

Place Table 2 about here

The correlation coefficients presented in Table 3 permit us to assess the similarity, or lack of similarity, of peer and student judgments concerning teaching effectiveness. Only the correlations between the subsequently discontinued, peer ratings of 1974 (all peers rate all other peers) and the one-term student ratings of 1974 are significant beyond the .01 level. This correlation, of .64 for all faculty and of .76 for experienced faculty, is somewhat higher than the .48 produced by Guthrie (1949), but well within the range of similar correlations presented in the other studies previously mentioned in an earlier section of this paper. The FEC 75 committee ratings and the 1975 student ratings produced a lower correlation of .34 for all teachers and only .20 for the more experienced teachers. However, the correlations involving the peer-nomination system, although all positive, are generally low and do not reach significance at the .05 level. The comparisons between the three-year summary evaluations produce a .28 correlation for all faculty and a .17 correlation for experienced faculty. Note: All correlations coefficients involving peer nomination-rankings have had their signs reversed since the larger peer-ranking scores really represent "low" ratings. An alternative measure can be derived from peer nomination-rankings, number of times nominated.

This measure neglects the rank information available in the chosen method but correlates quite well (r=.95+) with it.

Place Table 3 about here

The correlations between the student ratings of one year and the peer nomination-rankings of the following year are slightly higher than those involving the same year's data. However, these differences do not reach significance at the .05 level.

Stability of Peer Ratings

Peer ratings and peer nomination-rankings of teaching exhibit considerable stability from year to year. The correlations presented in Table 4 indicate that one-year stabilities for peer evaluations of teaching are .70, .78 and .82 for all faculty, and .75, .74 and .62 for the ratings of the more experienced faculty. The ratings of the Faculty Evaluation Committee (FEC 75) also correlate well with peer nomination-rankings for the previous year (.70, .78); well with peer nomination-rankings for the same year (.66, .68); and with nomination-rankings of the two subsequent years (.60, .61; .52, .57)--for all faculty and for only the experienced faculty. Peer evaluations of service and research are similarly consistent and produce cross-year consistencies which usually exceed .73.

Place Table 4 about here

An interesting note. Previous to 1974, members of the EDFI department used no common student ratings form. The Winter of 1974 was the first time that common student ratings data were produced for many faculty, and this data may have influenced subsequent peer evaluations. Table 5 presents the correlations between these 1974 student ratings and subsequent peer evaluations for experienced faculty. The 1974 peer ratings occurred approximately one month after the 1974 student data was gathered but before the student data was tabulated and returned to faculty. The correlations between the 1974 student ratings and peer evaluations are: .76 with 1974 peer ratings; .53 with peer nomination-rankings of 1976; and .33 (not significant) with 1977 peer nomination-rankings.

Place Table 5 about here

Peer Nominations, Activity Counts and Years of Service

When peers were asked to rank-nominate faculty as "Teachers," they were also asked to produce similar scores for "Service" and "Research," and three year (1975 to 1977) summary scores were similarly produced for the three functions. After the end of the 1976-1977 academic year, faculty were also asked to identify the seven, or fewer if necessary, departmental colleagues with whom they most associated--academically and/or socially--during the previous year. Twenty-two of the 26 faculty completed these "association" nominations. Three faculty were only nominated once, while one faculty member was nominated by 12 other faculty. The mean time nominated was 5.46, and the median was 5.

All faculty were asked to bring their vitae up to date by the end of the Spring of 1977 in order for the information to be included as part of a University self-evaluation. These vitae were examined to produce a count of books, book reviews, research reports and articles written, papers and workshops presented, speeches made, and grants received. Committee participation was also identified from the vitae, with one point accumulated for each year of service on any faculty committee and an additional two points added for each committee chairship held. All counts reflected only activities dated during the past three academic years. Service activities combined the workshop, speech and committee categories. Research Activities were defined as a weighted combination of books (x5), articles (x2) and papers, research reports, book reviews and grants (all x2). Years of service were determined as of June 1977.

Table 6 presents the correlations between these summary student ratings, peer nomination-rankings, peer association scores (number of times nominated), service activities (committees, workshops and speeches), research or scholarly production (books, articles, papers, research reports, book reviews, and grants), and years of service at BGSU. The largest correlation, of .84, was that between peer nomination-rankings for service and the count of service activities for the same three years. The comparable count for research activities was negligibly correlated (r=.16) with peer nomination-rankings of research.

The peer association measure was moderately correlated with: peer nomination-rankings of teaching (.48, p .05); peer nomination-rankings of research (.57, p .01); and the count of service activities (.54,

p .01). The only other correlations to reach a reasonable level of significance (p .05) were those: between peer nomination-ranking of teaching and peer nomination-rankings of research (.38); and between student ratings and years of service (.40).

Place Table 6 about here

Years of service appears most related to student ratings with the more experienced teachers eliciting somewhat better student ratings (r=.40). Peer evaluation of service was also positively correlated (r=.25) with years of service while negatively correlated (r=.25) with research activities, although neither correlation reaches significance at even the .1 level.

Potential Bias in Summary Peer Evaluations

The same peer nomination-ranking system was used during the past three years of departmental evaluation (1975 to 1977). Since the process asked faculty to rank up to seven peers for each of the three functions of teaching, research and service, it is possible that associations within groups of faculty could influence the nominations. It is evident that exactly the same faculty were not nominated in the same way for the three different functions, since the between-function correlations presented in Table 6 are relatively small: .12 between teaching and service; .38 between teaching and research; and .01 between service and research. However, the correlations between peer association and peer evaluations are larger: .24 with peer service, .55

with peer teaching, and .45 with peer research. The latter two correlations reach significance, and are greater than the correlations between the peer evaluations and factors more logically related to the functions: the corresponding correlation betwen peer and student evaluations of teaching is only .28; and the correlation between peer evaluation of research and a vita-derived measure of research is only .16. Manipulating the vita-derived research score so that books count less, or so that other activities are valued more, does increase the correlation to as high as .45 when all research activities are equally weighted. The small correlation between peer evaluation of service and association is not necessarily biased since there is a considerable correlation between the vita-derived service measure (mostly committee membership) and the association score (r=.54, significance=.002).

Table 7 presents the means and standard deviations for a wide range of variables potentially associated with student and peer evaluations while Table 8 presents the Pearson Product-Moment correlations between the four summary evaluations (mean student ratings and mean peer ranking-scores of teaching, research and service, for three years) and these variables. Some of these variables reflect the same three years reflected in the summary evaluations, the student credit hours (SCH) reflect the last four years, and a few of the variables reflect the total activity reported in the vita in order to determine if a history of activity influences the evaluation. The only consultantship score reflects the entire vita. Since many of the consultantships reported in the vita were not dated, there is no consultantship score for only the past three or four years. There is no measure of

number of programs or courses created since many faculty failed to report these activities in their vita even though this type of activity is valued by many members of the department.

Place Tables 7 and 8 about here

According to the data presented in Table 8, the three major predictors of student ratings are: years of service, number of undergraduate advisees and the percentage of the teacher's classes that are identified as graduate courses. The five best predictors of the summary peer evaluation of teaching are: visibility (or office location), a measure of peer association, the number of speeches and workshops presented during the past three years, the total number of student credit hours earned by teaching regularly scheduled classes, and peer evaluation of research. Neither years of service or percentage of graduate classes taught were strongly related to peer evaluation of research. Neither years of service or percentage of graduate classes taught were strongly related to peer evaluation of teaching even though they were related to student ratings. Student ratings were only slightly related to peer evaluations of teaching ($r = .28$, significance level = .11).

Office location/visibility and nomination as an associate were also related with peer evaluation of research, although less with peer evaluation of service. Peer evaluation of teaching and vita-counts of papers read, grants received, book reviews, and research reviews are the other two correlates of peer evaluation of research while committee membership is the major, and overwhelming correlate

of peer evaluation of service.

Multiple regression analysis was used by Aleamoni and Yimer (1973, pp. 276-277) to assess the contributions of various variables in the prediction of colleague ratings, and a similar procedure can be used in this case. However, the small number of cases requires us to view the results with considerable caution.

Peer service predictions. A stepwise, multiple regression analyses (of the only 22 cases with complete data) indicates that four non-evaluative variables produce a multiple correlation coefficient of .90 with summary peer rankings for the service function. The four variables, their associated "F" rations for removal/inclusion, and the simple correlations are: (a) committee membership (F=64.2, r=.85); (b) association nomination score (F=4.7, r=.24); number of speeches and workshops (F=4.18, r=.14); and number of books written (F=2.7, r=.09). When other peer evaluations and years of service are entered, the multiple squared correlation is increased from .81 to .96, primarily due to the addition of peer teaching evaluation.

Predicting peer evaluation of research. Peer evaluation of research is more difficult to predict through use of similar predictors. The first four variables which enter into a stepwise, multiple regression system are association score, speeches made, grants received and articles written, but the combined effect of these four variables only account for 33% of the peer research variance in comparison to 81% of variance of peer service predicted by the first (best) four predictors. Only the relationship between association score and peer-research produces an "F" ratio inclusion index over 1.5. The simple correlations

between peer-research and its logical predictors are: association (r=.43); paper presentations (r=.27); grants received (r=.27); speeches made (r=.20); and committee membership (r=.15). The substitution of total articles and papers-grants drawn from a longer than four year period increases the explained variance by only a few percent.

Predicting peer evaluation of teaching. Peer evaluation of teaching could be predicted through use of the four similar variables so as to include 47% of the common variance. Adding student evaluation of teaching increases the common variance to 72%. Only association score (r=.55), speeches made and workshops presented (r=.51), and student evaluation of teaching (r=.28) produce significant "F" ratios upon exit or entry. The other two variables which add about 6% of the common variance are: Number of undergraduate advisees (r=28) and SCHs for regular classes (r=.35).

Peer evaluations of teaching and research appear to overlap somewhat more than the two other possible peer-evaluation relationships. The simple correlation between peer evaluations of teaching and research (r=.38) just reaches significance at the .05 level. This relationship is the fourth best predictor of peer teaching (after association score, speeches and workshops, and student ratings of teaching) and the third best predictor of peer evaluation of research (after association score and numbers of papers/grants, etc.)

Other sources of bias in peer evaluation procedures. Several categorical variables may influence peer nomination-rank scores. Variables such as the centrality/visibility of office location, area affiliation and professorial rank can be investigated via a series of analyses

of variance with evaluation scores functioning as the criterion variables. None of these variables produce "F" ratios over 1.5 when student evaluations are the criterion variable. Office location (F=6.8, df=4, 15) is the only such categorical variable related to peer evaluation of teaching, and the trend is distinctly linear with the occupants of high visibility offices receiving better peer scores than the occupants of lower visibility offices. Neither office location (F=0.51, df= 4, 16), nor the other categorical variables is related to peer evaluation of service, and none are related to peer evaluation of research. Although the product-moment correlations between PR and office location are significant (r=.38, significance level=.04), tau is a more appropriate measure of relationship, and it is .19 (sign.=.14).

Potential Predictors of Student Evaluations of Teaching

The correlations presented in Table 8 indicate that neither peer evaluations nor other indicators of research or service are strongly related to student evaluation of teaching. Years of service at BGSU has, however, been shown to be correlated with student ratings (see Table 8; r=.40, level of significance=.031). Although this modest correlation could be attributed to chance, to teaching expertise gained through experience and practice, or to changing hiring practices, it could also be attributed to the greater power of the more experienced teacher to obtain smaller classes and/or graduate classes likely to produce somewhat higher ratings (see Centra and Creech, 1976, for evidence supporting this relationship between ratings and class characteristics). There is a positive correlation between the percentage of

small (15 or fewer students) classes taught and summary student ratings (r=.17, N=19, significance=.24), and between ratings and the percentage of graduate classes taught (r=.37, N=23, significance=.04), but neither of these factors is particularly related to years of service (r=-.03 and r=.15 for years of service vs. percent small class and percent graduate classes, respectively).

The total number of regular, on-campus classes taught in three years ranged from only 5 classes for one faculty member to 27 classes for nine faculty. The median number of classes taught in three years, or in nine academic quarters, was 22.2 with a mean of 20.4. Number of classes taught does correlate slightly and negatively with student ratings (r=-.21, N=23, significance=.17), but it does not correlate with years of experience (r=.01). However, the number of sections taught DOES correlate to some degree, and positively, with Peer Evaluation of teaching (r=.34, N=21, significance=.06). This correlation, although small, is somewhat larger than the correlation between student and peer evaluations of teaching (r=.28). Somewhat surprisingly there were no relationships between the indicator of load change and other variables. The rate of course change remained relatively constant over the four years investigated.

Conclusions and Discussion

The Eight Hypotheses: Five Supported and Three Unsupported

Previous investigations concerning faculty evaluation have produced several general patterns which were cast into the form of eight directional hypotheses.

1. Student ratings, especially for experienced faculty, produced a considerable degree of cross-year consistency, and correlations in the .7+ range are not atypical.

2. It was hypothesized that student and peer evaluations of teaching, even though based upon different definitions of teaching, would be correlated to a moderate degree. With the exception of the first year the system was employed, student and peer evaluations of teaching were not correlated, even though both evaluations produced moderate cross-year consistencies. The first year correlations could be attributed to the all-rate-all faculty evaluation system (which was subsequently voted down) or to a failure to define teaching during this first year. Both factors may force teachers to fall back upon the type of general student gossip likely to be an indirect reflection of student ratings. Office location/visibility, a faculty association measure, speeches and workshops counts, and numbers of on-campus students taught, all predict peer evaluation of teaching. Consequently, the peer nomination-ranking evaluation of teaching is clearly influenced by, and possibly biased by, faculty visibility-activity. Even when a faculty committee, similar to the committees reported by Guthrie (1949), operated in 1975, these FEC ratings of teaching were: correlated wth the peer nomination ratings; largely uncorrelated with student ratings for the same year; and as influenced by association as were peer-nomination rankings.

Without the use of independent criteria (such as measures of student learning, student satisfaction with advisement, or tests of teacher knowledge), it is difficult to characterize "visibility/activity"

as either bias or a legitimate correlate of the faculty teaching function. Regardless, it is quite apparent that neither student nor peer evaluation can be substituted for the other.

3. and 4. Student evaluations of teaching were not expected to be correlated with either peer evaluations of research or service or with other indicators of research and service. These predictions held. However, peer evaluations of teaching and research were correlated with one another, to some degree, as they have been in some previous studies.

5. Peer evaluations of service and peer evalutions of research were not correlated, although this correlation was expected due to a possible blurring of the service-research boundaries. Service was quite clearly associated with committee membership, and committee membership was quite characteristic of the majority of the departmental faculty since only two faculty were not members of committees, one seemed to average 12 committee memberships per year and the departmental mean committee membership exceeded three committees per year. Speeches and workshops were less related to service rankings, but were also less characteristic of the departmental faculty in general since ten faculty were credited with none, and the departmental mean total for a three year period only 2.4 (1.2 median) or less than one per year per faculty.

6. and 7. and 8. It was excpected that peer evaluations of teaching, research, and service, via a nomination-ranking system, would be correlated with a similar nomination-measure of faculty association. An extensive experience with "guddy rankings" in the military services has provided evidence that friendship or association measures influence peer nomination-evaluations even though this influence does not neces-

sarily invalidate the peer evaluations. A measure of peer-association was correlated with peer evaluation of teaching and peer evaluation of research and accounted for approximately 25% of the common variances. Peer-association was not, however, correlated with peer-evaluation of service, although a count of service activities, espeically committee memberships, was correlated with the peer-association measure.

Conclusions

The evaluation system used in the EDFI Department for the past three to four years produces relatively independent, stable, and reliable indicators of student evaluation of teaching and peer evaluations of teaching, research and service. The peer evaluation of service is quite clearly a reflection of faculty committee membership, but the peer evaluation of research is not as strongly associated with similarly obvious measures of research productivity--perhaps because a high degree of research activity has not been a major characteristic of the department. The prime indicator of effective teaching, student ratings, are not correlated with peer evaluations of teaching. Indicators of association or visibility or activity are quite useful in the prediction of the two peer evaluations which are not otherwise strongly correlated with more function-related measures--peer evaluations of research and peer evaluations of teaching. At this point, it is difficult to determine if such activity/visibility indicators are truly related to the teaching, or research, function or if they are primarily evidences of bias in the evaluation system.

The failure to predict peer evaluations of teaching or research through use of vita-derived indicators has been disappointing, but it doesn't necessarily prove that the peer evaluations are clearly erroneous. It is possible that some faculty earn higher peer evaluations through their willingness and ability to stimulate and/or assist their colleagues to function as better teachers or researchers. A colleague who frequently is willing and able to assist colleagues to computerize their data gathering or to improve their teaching may earn higher evaluations as a researcher or teacher, and also be more likely to be nominated as a regular associate, even though these activities are not reflected in either the vita or in student ratings of teaching. Whether or not tese other types of supportive activity are actually moderating peer evaluations, or whether these types of activity are legitimately characterized as "teaching" or "research," are still open to question.

As is true of any correlational study, and especially true of one which examines a large number of relationships which exist among a very limited number of people, any of the relationships that are apparently uncovered may be actually due to chance, to deosyncratic and atypical combinations which exist within this particular group but are not common to other groups--or they may reflect a realtiy which exists both within this department and also within others that are similar to it. Only a series of replications will permit one to separate the grain from the chaff. One recent, large-scale study (Jack E. Roose, August 1977) was a strong step in this direction, even though the purpose of the study was not to examine a department's evaluation system, but to use the methodology of "social judgment theory" and multiple regression

analysis to predict faculty rank and salary, given more than 28 vita-derived faculty characteristics and over 42 faculty judges.

Appendices

A copy of the Student Description of Teaching, an instructional evaluation form, and a copy of the suggested criteria for the peer-evaluations of teaching, research and service are attached to this paper.

References

Aleamoni, L. M., & Yimer, M. An investigation of the relationship between colleague rating, student rating, research productivity, and academic rank in rating instructional effectiveness. Journal of Educational Psychology, 1973, 64, 274-277.

Astin, A. W., & Lee, C. B. T. Current practices in the evaluation and training of college teachers. In C. B. T. Lee (Ed.), Improving college teaching. Washington, D. C.: American Council on Education, 1967.

Ballard, M., Rearden, J., & Nelson, L. Student and peer rating of faculty. Teaching of Psychology, April 1976, 3, 88-90.

Batista, E. E. The place of colleague evaluation in the appraisal of college teaching: A review of the literature. Research in Higher Education, 1976, 4, 257-271.

Bausell, R. B., Schwartz, S., & Purohit, A. An examination of the conditions under which various student rating parameters replicate across time. Journal of Educational Measurement, Winter 1975, 12, 273-280.

Bejar, I. I. A survey of selected administrative practices supporting student evaluation of instruction programs. Research in Higher Education, 1975, 3, 77-86.

Blackburn, R. T., & Clark, M. An assessment of faculty performance: Some correlates between administrator, colleague, student and self ratings. Ann Arbor: Center for the Study of Higher Education, University of Michigan, 1971 (mimeographed).

Bowling Green State University. Faculty charter. Unpublished manuscript, 1969.

Braunstein, D. N., & Benston, G. J. Student and department chairman views of the performance of university professors. Journal of Applied Psychology, 1973, 58, 244-249.

Bresler, J. B. Teaching effectiveness and governmental awards. Science, 1968, 160, 164-168.

Cartter, A. M. Ph.D's and the academic labor market. New York: McGraw-Hill Book Co., 1976.

Centra, J. A. Colleagues as raters of classroom instruction. Journal of Higher Education, 1975, 46, 327-337.

Centra, J. A. Faculty development practices in U.S. colleges and universities (ETS PR-76-30). Princeton, N. J.: Educational Testing Service, November 1976.

Centra, J. A. The student instructional report: Its development and uses (SIR Report #1). Princeton, N. J.: Educational Testing Service, 1972.

Centra, J. A. Two studies on the utility of student ratings for instructional improvement (SIR Report No. 2). Princeton, N. J.: Educational Testing Service, 1952.

Centra, J. A. The relationship between student and alumni ratings of teachers (Research Bulletin 73-39). Princeton, N. J.: Educational Testing Service, May 1973.

Centra, J. A., & Creech, F. R. The relationship between student, teacher, and course characteristics and student ratings of teacher effectiveness. In Two studies on the validity of the Student Instructional Report (SIR Report No. 4). Princeton, N. J.: Educational Testing Service, 1976, 17-74.

Clark, D. L., & Guba, E. G. RITE occasiona papers series. (Vol. 1-8). Bloomington, Indiana: Research on Institutions of Teacher Education, Indiana University, 1976.

Cope, R. G., McMillin, J. G., & Richardson, J. M. A study of the relationship between quality instruction as perceived by students and research productivity in academic departments. Washington: U. S. Office of Education, 1972.

Costin, F. Intercorrelations between students' and course chairmen's ratings of instructors. University of Illinois, Division of General Studies, 1966. (mimeographed)

Costin, F., Greenough, W. T., & Menges, R. J. Student ratings of college teaching: Reliability, validity, and usefulness. Review of Educational Research, 1971, 41, 511-535.

Crawford, P. L., & Bradshaw, H. L. Perception of characteristics of effective university teachers: A scaling analysis. Educational and Psychological Measurement, 1968, 28, 1079-1085.

Doyle, K. O. Faculty evaluation: Some considerations and a model. In A. L. Sockloff (Ed.), Proceedings of the first invitational conference on faculty effectiveness as evaluated by students. Philadelphia: Measurement and Research Center, Temple University, 1973, 59-90.

Doyle, K. O., Jr. Student evaluation of instruction. Lexington, Mass.: D. C. Heath & Company, 1975.

Dressel, Paul L. Student evaluation of faculty: Why? What? How? In A. L. Sockloff (Ed.), Proceedings of the first invitational conference on faculty effectiveness as evaluated by students. Philadelphia, Pennsylvania: The Measurement and Research Center of Temple University, April, 1973, 1-24.

Drucker, A. J., & Remmers, H. H. Do alumni and students differ in their attitudes toward instructors? Journal of Educational Psychology, 1951, 42(3), 129-142.

Faia, M. A. Teaching and research: Rapport or mesalliance. Research in Higher Education, 1976, 4, 235-246.

Faricy, W. H. Department affiliation and faculty reference group: Some effects on faculty opinions. Chapter 2 or P. L. Dressel & W.H. Faricy, Return to Responsibility, San Francisco: Jossey-Bass, 1972.

Feldman, K.A. The superior college teacher from the students' view. Research in Higher Education, 1976, 5, 243-288.

Fenker, R. M. The evaluation of university faculty and administrators: A case study. Journal of Higher Education, November/December 1975, 46, 665-686.

Fentress, J. H., & Swanson, R. A. The effect of instructor influential tactics on evaluation by university students. Journal of Industrial Teacher Education, 1975, 13(1), 5-16.

Francis, J. B. Faculty ratings of course evaluation items. Research in Higher Education, 1976, 4, 23-40.

French, G.M. College students' concept of effective teaching determined by an analysis of teacher ratings. Dissertation Abstracts, 1957, 17, 1380-1381.

Gaff, J. G. Teachers who make a difference. In A. L. Sockloff (Ed.), Proceedings of first invitational conference on faculty effectiveness as evaluated by students. Philadelphia, Pennsylvania: The Measurement and Research Center of Temple University, April 1973, 219-232.

Gillmore, G. M., Kane, M. T., & Naccarato, R. W. The teacher and the course as units of analysis in the generalizability of student ratings of instruction (EAC Report 77-9). Seattle, Washington: University of Washington Educational Assessment Center, February 1977.

Guthrie, E. R. The evaluation of teaching. The Educational Record, April 1949, 30, 109-115.

Guthrie, E. R. The evaluation of teaching: A progress report. Seattle: The University of Washington, 1954.

Hildebrand, M., Wilson, R. C., & Dienst, E. R. Evaluating University Teaching. Berkeley: Center for Research and Development in Higher Education, University of California, Berkeley, 1971.

Hogan, T. P. Similarity of student ratings across instructors, courses, and time. Research in Higher Education, 1973, 1, 149-154.

Hollander, E. P. The friendship factor in peer nominations. Personnel Psychology, 1956, 9, 435-447.

Hoyt, D. P. Measurement of instructional effectiveness. Research in Higher Education, 1973, 1, 367-378.

Hoyt, D. P., & Spangler, R. K. Faculty research involvement and instructional outcomes. Research in Higher Education, 1976, 4, 113-122.

Kemerer, F. R. The clouded future of faculty governance. The Educational Forum, January 1978, 43, 233-243.

Lester, R. Antibias regulations of universities: Faculty problems and their solutions. New York: McGraw-Hill Book Co., 1974.

Love, J. A., Minnis, D. L., Sandoval, J. H., & Yonge, G. D. An instrument to describe college teaching: A reliability study. Paper presented at the Annual Meeting of the American Educational Research Association, New York, N. Y., April 4-8, 1977.

Martin, W. B. (Ed.). Redefining service, research, and teaching. San Francisco: Jossey-Bass, Inc., Publishers, 1977.

Maslow, A. H., & Zimmerman, W. College teaching ability, scholarly activity and personality. The Journal of Educational Psychology, 1956, 47, 185-189.

Miller, R. I. Evaluating faculty performance. San Francisco: Jossey-Bass, 1972.

Morsh, J. E., Burgess, G. G., & Smith, P. N. Student achievement as a measure of instructor effectiveness. The Journal of Educational Psychology, 1956, 47, 79-88.

Murray, H. G. The validity of student ratings of teaching ability. Paper presented at the Canadian Psychological Association Annual Meeting, Montreal, 1972.

Roose, J. E. The use of social judgment theory to study the extent of unfair sex bias in the salaries of faculty members. Unpublished doctoral dissertation, Bowling Green State University, August, 1977.

Seldin, P., & Wakin, E. How deans evaluate teachers. Change, November 1974, 48-49.

Shull, H. The university tenure "Problem". Science, 8 October 1976, 194, editorial page.

Smart, J. C., & Montgomery, J. R. (Eds.) Examining departmental management. San Francisco: Jossey-Bass, Inc., 1976.

Sockloff, A. L. Instruments for student evaluation of faculty: Ideal and actual. In A. L. Sockloff (Ed.), Proceedings of the first invitational conference on faculty effectiveness as evaluated by students. Philadelphia: Measurement and Research Center, Temple University, 1973, 132-151.

Spencer, R. E. A history of the development of the Illinois course evaluation questionnaire (Research Report #306). Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1969(a).

Spencer, R. E. Research report: Some dimensions of the Illinois Course Evaluation Questionnaire (Research Report #303). Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1969(b).

Swanson, R. A., & Sisson, D. J. The development, evaluation, and utilization of a department faculty appraisal system. Journal of Industrial Education, 1971, 9(1), 64-79.

Webb, W. B., & Nolan, C. Y. Student, supervisor, and self-ratings of instructional proficiency. The Journal of Educational Psychology, 1955, 46, 42-46.

Wherry, R. J., & Fryer, D. H. Buddy ratings: Popularity contest or leadership criteria? Personnel Psychology, 1945, 2, 147-159.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Wolins, L. The validity of peer ratings as a function of rater characteristics and methods of scoring. Unpublished doctoral dissertation, Ohio State University, 1956.

APPENDICES

STUDENT DESCRIPTION OF TEACHING

(used by EDFI, Firelands Campus, HOEC, IET, and faculty in EDAS, EDCI and EDSE

We hope to improve teaching by asking you to describe your teacher's performance -- and to suggest improvements. Results will be given to your teacher after the grades have been turned in. Some departments use this data for administrative decisions.

This form should be presented at the beginning of a class period -- with the teacher absent from the room. Mark the IBM sheet with a soft pencil. A student should return all forms to Room 455 Education.

Please write the teacher's name and course and section number on the IBM sheet.

USE THESE NUMBERS TO DESCRIBE THE INSTRUCTOR

1- Below average
2- Average
3- Above average (Best 50%)
4- Better than most (Best 30%)
5- Superior (Top 10% - 15%)
6- One of the best - equal to any previous teacher
7- The best - better than any previous teacher

Please mark your answers (in pencil) on the IBM sheet.

1. SCHOLARSHIP (Has command of the subject, contrasts various points of view, discusses recent developments in the field, presents origins of ideas and concepts)

2. ORGANIZATION/CLARITY (Explains clearly, is well prepared, presents material in an organized manner.)

3. INTERACTION WITH GROUP (Is sensitive to the reactions of the class, encourages student participation, welcomes questions and discussions.)

4. INTERACTION WITH INDIVIDUALS (Has genuine interest in students relates to students as individuals, respects students as persons, is valued for his advice.)

5. ENTHUSIASM (Seems to enjoy teaching, is a dynamic and energetic person, has an interesting style of presentation, is enthusiastic about the subject, has a sense of humor.)

Using the same scale, provide an overall rating for the following.

6. THE INSTRUCTOR

7. THE TOTAL EXPERIENCE UNDER THE CONTROL OF THIS PERSON (Teaching, content, readings, tests, outside activities, seminars, homework, . . . everything that you associate with this class or lab/section/seminar/group.)

8. STUDENT ACCOMPLISHMENT (The degree to which this class helped you to increase your knowledge, skills and competencies in comparison with other classes you have taken.)

---

9. GRADE THIS COURSE IN COMPARISON TO YOUR OTHER COURSES THIS TERM:
   1) A 2) B 3) C 4) D 5) F

10. WHAT GRADE DO YOU EXPECT TO RECEIVE IN THIS COURSE:
    1) A 2) B 3) C 4) D,N or U 5) F

11. WHAT IS YOUR BGSU CUMULATIVE AVERAGE?
    1) 3.50 - 4.00 2) 3.00 - 3.50 3) 2.50 - 3.00 4) 2.00 - 2.50 5) Below 2.00

12. WAS THE INSTRUCTOR IN THE CLASSROOM WHEN YOU COMPLETED THIS FORM:
    1) Yes
    2) No

(Questions 1-5 were adapted from the Berkeley scales.)

On the back of the IBM sheet, please inform your instructor as to (1) which class activities, presentations, assignments and classes were the most valuable for you? (2) which were of lesser value? (3) how could this instructor have helped you to learn more effectively.

Regarding this course:

13 The difficulty level was:
- 1- Very elementary
- 2-Somewhat elementary
- 3- About right
- 4- Somewhat difficult
- 5- Very difficult

14 The work load was:
- 1- Very light
- 2- Somewhat light
- 3- About right
- 4- Somewhat heavy
- 5- Heavy

15 The course objectives were:
- 1- Very clear from the beginning
- 2- Very clear (eventually)
- 3- Moderately clear
- 4- Not too clear
- 5- Unclear

Use the following code to describe your attitude toward this course: 1. Definitely false 2. More false than true 3. Neutral - ?? 4. More true than false 5. Definitely true

16 I had a strong desire to take this course.
17 I would like to take another course from this instructor.
18 I would like to learn more about this field.
19 The course content was not related to my current or future interests.

The following questions are intended to provide the teacher with more SPECIFIC information - especially concerning areas which could be improved in future terms.

USE THESE NUMBERS TO DESCRIBE THE INSTRUCTOR

0 Omit (not relevant)

1 Major improvements needed (generally poor now)
2 Improvement is needed(is mediocre now)
3 Little improvement is needed (well done now)
4 No improvement is needed (already great)
5 Extraordinarily effective now

GRADING PROCEDURES
- 20 Fairness (so far)
- 21 Quality of tests
- 22 Fairness in grading of assignments
- 23 Clear explanation of how student performance is to be evaluated

TEACHER'S ABILITY TO
- 24 Speak clearly
- 25 Lecture
- 26 Stimulate discussion
- 27 Motivate students to learn
- 28 Individualized instruction to provide useful options for different students
- 29 Inspire excitement and interest in course content
- 30 Ask thought-provoking questions

ACTIVITIES AND MATERIALS
- 31 Textbooks
- 31 Supplementary readings
- 33 Movies and use of other AV devices
- 34 Effective integration of class-work and assignments

INSTRUCTOR
- 35 Sense of humor
- 36 Preparation for class
- 37 Openness to student questions
- 38 Respect for students

ORGANIZATION
- 39 Clarity of objectives
- 40 Clarity of presentations
- 41 Stress on important areas
- 42 Summarization of major points
- 43 Effective use of class time

(by P. Wood)

STUDENT RATINGS OF IN-CLASS TEACHING

A student ratings questionnaire, STUDENT DESCRIPTION OF TEACHING, has been developed to elicit student opinions concerning faculty and course characteristics. The first five questions reflect those teacher characteristics frequently identified as being valued by students. The students are asked to use a seven point scale to indicate the degree to which their instructor possesses these characteristics (1=below average, 2=average, 6=equal to any previous teacher). The next three questions (6, 7 and 8) ask the student to use the same seven point scale to rate: the instructor; the total experience; and student accomplishment. These three responses are taken as representing the students' overall evaluation of the teacher and the course. The mean student response to these three questions--elicited from all students, from all regularly-assigned, on-campus, non-extension classes taught during the spring quarter and the following fall and winter quarters--is calculated to produce a single student ratings score for the year. The remaining questions of the form are intended to produce data of possible interest to faculty, and their use is entirely optional.

Administration: All departmental faculty are asked to present the ratings forms to all on-campus classes at the beginning of the last class period in each term. The students are to pencil-mark the associated IBM 555 optical scan form while the instructor is absent from the room. A student volunteer is to gather the forms, replace them in the class envelope and return them by hand or campus mail to a departmental office (Room 529 or 455). All original and tabulated data is returned to the faculty member to be used as the faculty member wishes-although the single student ratings score is also available to the Faculty Evaluations Committee and will be considered in the development of personnel decisions.

Bias: Although it is possible-even likely--that some types of classes regularly produce higher ratings than other types of classes, these possible biases may be corrected through a negotiation with the departmental chairperson to obtain a fairer assignment of classes--or through calling the possible biases to the attention of relevant personnel committees.

Reliability: When used with classes of more than ten to fifteen students, each of the three relevant questions produces an analysis of variance-produced estimate of reliability that is quite high (.85 to .93). Although there is little evidence that students are completely capable of evaluating all aspects of faculty performance, there is no reason to believe that students can not provide an estimate of their perception of the teacher's effectiveness, of the total course experience, and of their own learning. The high within-class consistency of most ratings questions tends to support this belief-as does approximately 40 years of extensive ratings use at a variety of colleges.

sshl/26

TEACHING

1) Effectiveness in stimulating students to learn
2) Knowledge of content area
3) Effectiveness in sharing teaching competencies with colleagues
4) Efforts to improve teaching effectiveness
5) Effectiveness in advisement
6) Supervision of thesis, dissertation, and/or independent study
7) Development of innovative courses or programs.

SCHOLARLY OR CREATIVE EFFORTS
(Publications, Programs, Research)

1) Has the necessary competencies to produce scholarly or creative efforts
2) Develops proposals, publications, papers, programs, presentations
3) Significant in influence on faculty, organizations, school systems, programs . . .
4) Improves the quality and quantity of scholarly creative efforts by interaction with other faculty, attendance at workshops or professional meetings, extensive reading . . .
5) Functions as a consultant

SERVICE

1) Is an active and valuable contributor to university committees or groups (at the area, department, college, and/or university or state level)
2) Provides service to peers and colleagues
3) Is an active and valuable contributor to professional associations
4) Provides professional public service beyond this campus to assist other universities, colleges, schools, agencies, companies--(not including "good citizenship" activities performed in the capacity of a concerned citizen in church, youth groups, etc.)
5) Has received special awards and/or recognition for professional service.

A

Bowling Green State University

Department of Educational
Foundations & Inquiry
Bowling Green, Ohio 43403
(419) 372-0151 ext. 322

April 29, 1975

MEMORANDUM

TO: EDFI Faculty

FROM: EDFI Faculty Evaluation Committee

RE: Peer Input to Faculty Evaluation

After considerable discussion, the EDFI Faculty Evaluation Committee has decided that peer input is an important--and unique--source of information relative to decisions concerning EDFI faculty. The peer input procedure used last year--everyone rate everyone--has too many obvious logical and psychometric disadvantages. The system in which each faculty member asks several peers to provide ratings-recommendations also has many disadvantages. A third approach combines simplicity with psychometric and logical reasonableness--while still producing a type of peer opinion likely to be a valuable supplement to student ratings, chairperson ratings and committee opinion.

Our department is so large and its members' interests and accomplishments are so diverse that it is unreasonable to believe that all of us are aware of the contributions and strengths of all members. However, our faculty are making valuable contributions in the areas of TEACHING and/or RESEARCH and/or SERVICE, and these contributions--many of which are not adequately represented in vita or known to all of us--are known to some of their colleagues. This knowledge can be transmuted into input to FEC decisions via the following colleague-perception-of-contribution system. Each of the following three pages provides a set of questions and/or statements which partially define one of the three areas of academic contribution--Teaching, Research/Production, and Service. Each page also contains a list of EDFI faculty. Faculty members are asked to:

(1) decide upon their own definition of teaching (or research or service);
(2) indicate which faculty member--to their knowledge--best exemplified this definition during the past year;
(3) indicate this person by writing the number "1" in the space next to that person's name;
(4) indicate who is the next best exemplar of their definition by placing a number "2"; and
(5) continue this procedure until a minimum of two faculty and a maximum number of seven faculty are ranked on each of the three areas of contribution (faculty, of course, may nominate themselves in the position that they consider most appropriate).

Faculty resumes for most (many) faculty are available in the departmental office for those who wish to view them. Please return these forms to Cathy Long next week -- May 5 to May 9.

TABLES

Table 1

Means, Standard Deviations and Correlations: Student Ratings: 1974 to 1977

Experienced Faculty			All Faculty				Correlations			
N	Mean[a]	SD	N	Mean	SD	Year	1974	1974	1976	1977
16	4.26	.66	19	4.25	.62	1974[b]	—	.65**	.64**	.54*
20	4.33	.51	26	4.13	.65	1975	.49*[c]	—	.71***	.79***
19	4.36	.57	25	4.28	.66	1976	.46*	.75***	—	.81***
17	4.21	.58	23	4.28	.65	1977	.38	.45*	.72***	—

Notes: [a]Ratings represent a 1 to 7 point scale with 1=low and 7=excellent.
[b]Ratings for 1974 represent only one academic quarter, others represent three quarters each.
[c]Ratings below the diagonal represent all faculty; those above, faculty hired before 1972.
*p<.05
**p<.01
***p<.001

Table 2

Peer and Student Evaluations of Teaching, 1974 to 1977
Means and Standard Deviations for Experienced, and for All, Faculty

For All Faculty						Year	For Experienced Faculty					
Peer			Student				Peer			Student		
N	Mean	SD	N	Mean	SD		N	Mean	SD	N	Mean	SD
23	2.8	.52	19	4.24	.62	1974	20	2.9	.54	16	4.26	.66
26	189	18	26	4.13	.65	1975	20	189	.57	20	4.32	.51
26	16.3	3.0	--	----	---	FEC 75	20	16.7	3.0	--	----	---
25	218	19	25	4.28	.66	1976	19	220	15	19	4.36	.57
25	155	20	23	4.28	.65	1977	19	157	15	17	4.21	.58
21	184	17	23	4.24	.59	Sum of 1975-1977	16	185	14	17	4.32	.52

Table 3

Correlations Between Peer and Student Evaluations of Teaching
1974-1977

Comparison	Correlation Coefficients, numbers and significance levels					
	All faculty			Experienced faculty		
	r	N	Sig.	r	N	Sig.
1. 1974 Peer ratings - Student ratings	.64	19	.002	.76	16	.001
2. 1975 Peer nominations - Student ratings	.24	26	.115	.03	20	.447
3. 1975 FEC ratings - Student ratings	.34	26	.045	.20	20	.203
4. 1976 Peer nominations - Student ratings	.33	24	.060	.17	18	.251
5. 1977 Peer nominations - Student ratings	.28	23	.099	.16	17	.271
6. 1975 Peer - 1974 Student	.44	19	.027	.53	16	.018
7. 1976 Peer - 1975 Student	.39	25.	.029	.36	19	.066
8. 1977 Peer - 1976 Student	.37	24	.036	.25	18	.154
9. Sum all student ratings vs. sum all peer nominations, 1975-1977	.28	20	.113	.17	15	.277

NOTE: The signs for all correlations involving peer nominations have been reversed since a low peer nomination score really indicates a "high" rating.

Table 4

Peer Ratings/Rankings of Teaching: 1974 to 1975

Ratings	1974	1975	FEC75	1976	1977
1974	---	.75***	.68***	.48*	.68***
1975	.70***	---	.68***	.74***	.75**
FEC75	.70***	.66***	---	.61**	.57**
1976	.43*	.78***	.60***	---	.62**
1977	.53**	.75***	.52**	.82***	---

Notes: Correlations above the diagonal represent faculty hired before 1972, while correlations below the diagonals represent all faculty.

*p<.05

**p<.01

***p<.001

Table 5

Correlations Between Original Student Ratings (1974)
and Subsequent Peer Evaluations (for Experienced Faculty)

Peer Evaluation	N	Mean	SD	Correlation with Student 74	N	Significance
Peer Ratings 1974	20	2.9	.66	.76	16	.001
Peer Ranking 1975	20	189	18	.53	16	.018
Peer Ranking 1976	19	220	15	.60	15	.009
Peer Ranking 1977	19	157	15	.33	15	.117

Table 6

Means, Standard Deviations and Correlations For: Student Ratings, Peer Rankings, Service and Research Activity Counts, Peer Association and Years of Service Summary Data for 1975 to 1977

Variables	Cases	Mean	SD	Correlations 1	2	3	4	5	6	7	8
1. Student Ratings	23	4.24	.59	--	28	-.24	-.10	-.21	-.07	.02	.40*
2. Peer/Teaching	21	184	18	20	--	.12	.38*	.26	-.14	.55**	.06
3. Peer/Service	22	185	23	21	19	--	.01	.84***	-.30	.24	.25
4. Peer/Research	23	186	14	22	21	21	—	.24	.16	.45*	-.12
5. Service Activities	26	12.5	8.3	23	21	22	23	--	.16	.54**	.03
6. Research Activities	26	9.8	9.9	23	21	22	23	26	--	-.08	-.25
7. Peer/Association	26	4.0	2.3	23	21	22	23	26	26	--	-.03
8. Years of Service	26	9.5	5.1	23	21	22	23	26	26		--

Notes: Number of cases are placed below the main diagonal.
*p<.05
**p<.01
***p<.001

Table 7

Means, Standard Deviations (SD) and Number of Cases (N) for Several Possible Predictors of Student and/or Peer Evaluations: Most Data Drawn from Faculty Vitae

Variables	Mean	SD	N
1. Association=times nominated as close associate	5.46	3.2	26
2. U. Graduate=typical number of undergraduate advisees	18.4	18.6	26
3. Graduate=typical number of graduate advisees	18.2	30.2	26
4. Dissertation=total number of theses committees	4.7	8.3	26
5. Articles=total number of "published" articles	4.8	7.0	26
6. Articles: 3 yrs.="published" articles: 3 yrs.	1.6	3.2	26
7. Papers=all papers, book reviews, grants, reports	10.2	7.5	26
8. Papers: 3 yrs.=papers, etc.: last 3 yrs.	4.0	4.6	26
9. Speeches=speeches and workshops reported on vita	6.6	7.5	26
10. Speeches: 3 yrs.=speeches, workshops: 4 yrs.	2.3	3.0	26
11. Consultations=all consultantships reported on vita	7.4	8.2	26
12. SCH=student credit hours, regular classes: 3 yrs.	2847	1428	26
13. Extra SCH=SCH for other activities: 4 yrs.	413	646	26
14. % Small=% classes with 1-15 students: 4 yrs.	23%	20%	26
15. % Graduate=% graduate classes: 4 yrs.	41%	39%	26
16. Years Service=Years at BGSU as of 1977	9.5	5.0	26
17. Committees (1)=Committee membership: 3 yrs.	10.0	8.8	26
18. Visibility=Centrality of Office Location	2.7	1.2	25
19. S.T.=mean student rating: last 3 yrs.	4.27	.59	23
20. P.T.=peer ranking score: 3 yrs.: Teaching	184	18	21
21. P.R.=peer ranking score: 3 yrs.: Research	186	14	23
22. P.S.=peer ranking score: 3 yrs.: Service	185	23	22

NOTES: (1) Committees=(membership or departmental, college or university committees) x (years of membership since 1975) + (2) (committee chairships since 1975).

Table 8

Correlations between Student Evaluation of Teaching (ST), Peer Evaluations of Teaching (PT), Research (PR) and Service (PS) AND Possible Predictors of Evaluation

Variables	Correlations with ST	PT[1]	PR	PS
1. Association	.12	.55**	.45*	.24
2. U. Graduate	.39*	.33	.07	-.03
3. Graduate	.25	-.10	.14	.05
4. Dissertations	.26	-.03	.13	.00
5. Articles	.15	-.25	.23	-.27
6. Articles: 3 yrs.	-.19	-.09	.11	-.23
7. Papers	.06	-.25	.35*	-.09
8. Papers: 3 yrs.	-.05	.13	.42*	-.18
9. Speeches	-.08	.44*	-.08	.14
10. Speeches; 3 yrs.	.08	.51**	.14	.15
11. Consultations	.21	.12	.19	-.09
12. SCH's	.04	.42*	.24	-.16
13. Extra SCH's	.20	.24	.07	-.08
14. % Small	.17	-.19	-.29	.42*
15. % Graduate	.37*	-.22	-.19	.17
16. Years Service	.40*	.06	-.13	.25
17. Committees	-.23	.06	.17	.84***
18. Visibility[2]	.04	.62**	.38*	.23
19. S.T.	--	.28	-.10	-.24
20. P.T.	.28	--	.38*	-.12
21. P.R.	.10	.38*	--	.01
22. P.S.	-.24	.12	-.01	--

Notes: [1] The signs for all correlations involving peer nomination-scores have been reversed since a low peer nomination score indicates a "high" rating.

[2] Kendall's tau correlationship measure is a more appropriate measure of this relationship and it is: -.01 with ST; .52 (p .002) with PT; .19 (p .14) with PR; and .22 (p .11) with PS.

*p .05
**p .01
***p .001